A METHOD FOR STORING INFORMATION IN DNA

Field of the invention

The present invention relates to a method for storing information in DNA. The method of invention comprises storing information in DNA. The present invention addresses storage for all kind of digital information whether it is a text file, an image file or an audio file. Large sequences are divided into multiple segments.

Background of the invention

DNA is the best molecular electronic device ever produced on the earth because DNA can store, process and provide information for growth and maintenance of living system. All living species are as a result of single cell produced during reproduction. In most of the cases this single cell does not have most of the materials required for fabricating a living system but contains all the information and processing capability to fabricate living spaces by taking materials from environment, for example, fabrication of baby from Zygote which contains rearranged DNA sequences of parents. DNA is ready to use nanowire of 2 nm and can be synthesized in any sequence of four bases i.e. ATGC. DNA of every living organism (micro/macro) consist of large number of DNA segments where each segment represents a processor to execute a particular biological process for growth and maintaining life. Other important characteristics of DNA which makes it material of choice for future molecular devices are: DNA the building block of life, can store information for billion of years, . The tremendous information storage capacity of DNA can be imagined from the fact that 1 gram of DNA contains as much information as 1 trillion CD's[1,] four bases (A,T,G,C) instead of 0 and 1, extremely energy efficient (10^{19} operations per joule), synthesis of any imaginable sequence is possible and semiconductor are approaching limit.

Clelland et al., 1999[2], and Bancroft, et al. 2001[3] [U.S. patent no. 6,312,911], have developed the DNA based steganographic technique for sending the secret messages. Although their prime objective was steganography (the art of information hiding), they used DNA as storage an transmission device for secret message. They encrypted the plaintext message into the DNA sequences and retrieved the message using the encryption/decryption key. They used three DNA bases for representing a single alphanumeric character, as DNA has 4 bases (A, T, C, G) so a maximum of 64 (4x4x4) ASCII character can be formed using this scheme. Whereas, a total of 256 extended ASCII characters are required to represent complete set of digital information. Hence, Clelland's scheme cannot be used to address complete set of digital information and has limited scope.

EXPRESS MAIL LABEL
NO.: EV 481667504 US

Objects of the invention

The main object of the present invention is to develop a comprehensive DNA based information storage technique.

Another object of the present invention is to encrypt complete extended ASCII character set in terms of minimum number of DNA bases.

Another object of the present invention is to develop software to encrypt/decrypt data in terms DNA bases.

Yet another object of the present invention is to design suitable primers to be flanked at both ends of the encrypted and synthesized information.

Summary of the invention

The present invention provides a method for storing information in DNA. The method of invention comprises storing information in DNA. The present invention addresses storage for all kind of digital information whether it is a text file, an image file or an audio file. Large sequences are divided into multiple segments

Brief description of the accompanying drawings

Fig. 1a.	Information storage in DNA. Structure of prototypical single segment information storage in DNA strand.
Fig. 1b.	Information storage in DNA. Structure of prototypical multi segment information storage in DNA strand.
Fig. 2.	Encryption of extended ASCII character set in terms of DNA bases
Fig.3.	Encryption Key. Extended ASCII characters in terms of DNA strands
Fig.4.	Process sheet for encryption & storage
Fig.5.	Process summary

Detailed description of the invention

The present invention provides a method for storing information in DNA. The method of invention comprises storing information in DNA. The present invention addresses storage for all kind of digital information whether it is a text file, an image file or an audio file. Large sequences are divided into multiple segments

The method enables the storage of information in DNA. In another embodiment a software based on the above method enables all 256 Extended ASCII characters to be defined in terms of DNA sequences. The basic concept used is to take minimum number of bases to define each Extended ASCII character. With simple permutation we have 4 sequences combinations with one base i.e. A, T, G, C. Similarly, with 2 bases we have 4x4=16 different sequences, with three bases we get 4x4x4=64 distinct sequences and four bases give

4x4x4x4=256 distinct sequences. Therefore, with a set of 4 bases, complete extended ASCII set has been encoded. Software named as "DNASTORE" has been developed in Visual Basic 6.0 for encryption and decryption of digital information in terms of DNA bases. Using DNASTORE complete extended ASCII character set can be encoded 256 different ways.

In yet another embodiment in our scheme, plain text/image or any digital information is encrypted in terms of DNA sequences using encryption key (software). If the information overflows the limits i.e. it cannot be synthesized in a single piece then it is encrypted and fragmented in a number of segments. Synthesis of encrypted sequence(s) is carried out using DNA synthesizer.

In yet another embodiment a fixed number of different DNA primers sequence have been designed and assigned a number, which resembles the segment position it represents e.g. segment 1, segment 2..... segment n. These are called as header primers. Two tail primers have also been designed one resembles continuation and other resembles termination segment.

In yet another embodiment the DNA segment(s) is/are flanked by known PCR primers [as described earlier] at both the ends i.e. header primers are attached at the beginning of segment and tail primers are attached at the end of the segment. If there is only one segment, at the beginning it is flanked by header primer number 1 and at the end it is flanked by termination tail primer. However, if there are more than one segments, each segment would be attached with header primers numbered as 1, 2,3...n respectively, at the end these would be attached with a continuation tail primer except for last segment which would be attached with a termination tail primer.

The SM DNA is then mixed with the enormous complex denatured DNA strands of genomic DNA of human or other organism. As the human genome contains about $3x10^9$ nucleotide pairs, fragmented & denatured human DNA provides a very complex background for storing the encrypted DNA. The DNA can be stored and transported on paper, cloths, buttons etc.

In still another embodiment only a recipient knowing the sequences of both the primers [starting and tail] would be able to extract the message, using PCR to isolate & amplify the encrypted DNA strand. Isolated and amplified DNA can then be sequenced using automated DNA sequencer. The DNA sequence obtained can then be converted into digital message using encryption/decryption key (software key).

In yet another embodiment the key is helpful in the secret & secure transfer of information particularly for spying and military purposes. It may also be helpful in anti-theft, anti- counterfeiting, product authentication, copyright infringements etc.

Table 1. Comparison of present art with existing art

S. No.	**Existing art** Clelland et al., Bancroft, et al.	**Reported invention**
1.	Uses unique 3-base sequence for each alphanumeric character	Uses unique 4-base sequence for each alphanumeric character
2.	Can represent a maximum of 64 (4x4x4) characters	Can represent a maximum of 256 (4x4x4x4) characters
3.	Can represent only 1/4th of extended ASCII character set	Can represent complete extended ASCII character set
4.	Cannot be used encrypt complete digital information i.e. meant for alphanumeric characters only	Can be used encrypt complete digital information as shown in examples

Example 1. Encryption and decryption of a textual message "CSIR" in terms of DNA bases may be defined as

a) Generation of an array of 256 elements (unique 4-base per character i.e. ATGC, ATGA, ATGT, ATGG). These elements represent complete extended ASCII character set values.

b) The input information is then encrypted character-by-character using array generated in step 1. The basis is ASCII values of each character is matched with the element no. of the array of step 1.

Encryption of the text "CSIR" in terms of DNA bases may be: **<u>TATGTTTCTATTTTAC</u>** where:

C is represented by DNA sequence TATG
S is represented by DNA sequence TTTC
I is represented by DNA sequence TATT
R is represented by DNA sequence TTAC

c) If the information overflows the limits i.e. it cannot be synthesized in a single piece or because of any other problem, then the encrypted sequence is fragmented in a number segments.

d) Encrypted segment(s) is/are then flanked on each side with header and tail primers.

e) Synthesis of encrypted sequence(s) is then carried out using DNA synthesizer.

f) The synthesized DNA segment(s) is/are then be kept separately or can be mixed up with the enormous complex denatured DNA strands of genomic DNA of human or other organism. As the human genome contains about $3x10^9$ nucleotide pairs, fragmented & denatured human DNA provides a very complex background for storing encrypted DNA.

g) The encrypted DNA can then be transported on paper, cloths, buttons or through any other medium.

Isolation decryption of above encrypted DNA sequence **TATGTTTCTATTTTAC** :

a) Isolation and amplification of encrypted DNA is done using known primers flanked at each end by PCR method.

b) Retrieved SM DNA is sequenced using DNA sequencer.

c) Obtained sequence is interpreted (integrated if multi-segment before interpretation) using DNASTORE software. The basis for retrieval is a string of 4-bases each at a time is taken and matched with array as generated in step 1 of encryption and storage. The element number of matching value is taken and converted to its ASCII equivalent

If the retrieved sequence is TATGTTTCTATTTTAC. The Decryption would be:

first 4-bases i.e. "TATG" would be in the array storage and encryption 67 = C

next 4-bases i.e. "TTTC" would be in the array of storage and encryption 83 = S

next 4-bases i.e. "TATT" would be in the array storage and encryption 73 = I

next 4-bases i.e. "TTAC" would be in the array of encryption 67 = R

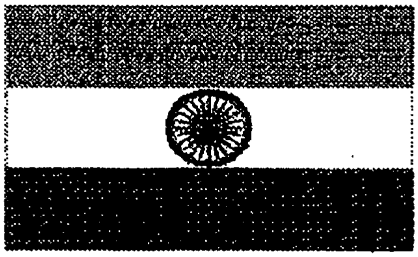
Integration of above decrypted values in the same sequence as retrieved is "CSIR".

Example 2. Some examples of DNA encryption for textual data

Digital Information	**Encrypted DNA sequence**
WELCOME	TTAGTACATAGCTATGTACCTAACTACA
WORLD PEACE	TTAGTACCTTACTAGCTATAAGCTTTCCTACATAGG TATGTACA
INDIA	TATTTATCTATATATTTAGG
CSIR	TATGTTTCTATTTTAC
CSIO	TATGTTTCTATTTACC

Example 3. A JPEG image encrypted in term of DNA bases

Digital Information	**Encrypted DNA sequence**
	TAAATATTTAGAAAACAATCTCGTGGCGATCGCGC CATCGGCTAACCTATCGATCGCTGGTCGCGTATCAA CAATCGTCGGTCGGTCGCGCCCTACGGGCTCTTCGA ACCCCGTAGGCGACACGGCGCGGCGGATGATTGTC GCCTTGCTACCCGTGGTGCGCCCAGACCTTCGACGC TCCTGGTACCTGCGCCTCATCGTTATCTTTGTTGGA GTGCAAGATGGAGAGTTTCCCGGACGGGTAGCAAG CCTGCGTAATATCTCCAAATGTCCAAAGCTTATTGT TTTCAATAACGTGATCCTTTACCTGCACATTAGTAT TATCACCAGCGTGCACCCATGCGGGCGCCAACCTT GCTGGACTTCGACGCCGCTGTCGTTGCCCTCTGAGT GAATGATTGTGCCCACTGTGGTGGGGCGCCTAGTC GGTCGGTCGAGGTGTTCATTAATGGATCGATCGAC



CTATCGAGGAATCGATCGATCGATCGGGCGATCGC
GCCATCGATCGATCAGTCGTCCTACGCCGGCTCTCT
CTGCATTTCAGCTCGCTTATCGAGAGGCCTGTGCAA
GGAGCCCTGTTACATTGGGCTATCTAAGACATGGG
GACAGTCGGCCGACAGAGTATAATAGGAACCACGC
CTAATGGATAACAGCTTTCGAAACCCACTCCAGAG
CCTGTTTACTCTAATTGGCTCCGGGGCTGATGGTGA
GGGCTGTGAACCCGGACTCCCAGCCTAGGGAGTAC
AGACCATGATCCCTATGCCGGATTAGCCCTAGGCT
GTCACACTAAGCTATCCTCAGCGTGAGCGTGTCCG
GACTTCGCAGGCTGTGCGTCTTGAGTGCGCGAGTG
GACGGGCGTGCGGATCCGCGCACGAACGCTTCGTC
GTTCGGTCGTCTTCACGACCGCCCAACTTTCCAGCC
ATCCAGGTAGCCACGCAAGCACATACACATACAGA
CATTTTATAATCCACTCTATTATCCAATCTTTCTGCT
GATCTGTCTACCTCGTAGGCTCCCTGGCTTAAGTGC
TAACTCACCAAAGTCCCGACCTACCAACCCTCCGTC
TTACCACCCTCCTCGCCGCCCGGCTGCCCTGCCCGC
TATGCGGGCAGCATTGCTAGCCACACAGCAAGCAT
CAGGGCCTGCGTCAACGCACGCTCCGTCGGCCGGG
CCGCTGGTCGGTGCGGAGGGGGGAGCGAGGGTAG
GCATGTGGGGTGGATCGCGCTTGGACTCCTCGGCT
GATTTGCTGACCGAGCCGTAGAATGATGCTCAGAA
GGAGATCGAGATAGACACGATACTTATCAGTCTGT
GTGTATGTACGTTCGTCCGTGCGTGGGTAGGTTGGT
CGATCGATTGATCTACGTTAATCCCACTCTGCGGCG
TGACATAATGAATTACCCGCCGCCCACTGTGCTGCG
AAACCCAGTTTACTCAGTTAATCCGACTATGCCACG
GTACAAAATATCCGGGGTGCATCCGACTTTGCAAA
TGAATCTAAAGCGCTACGTTATTGTAAAGATCGTA
ATTAACGAAGCGGTCGTTAATTAATCTGAGGTGCA
GATGAATACATTTAAACCATGCAGTTATTCATCAGT
CGCATCGCAAACTTGTAGACGCTGAATATTAGGTA
TGATTAATGATACGCGTGATGACAATTACGTGTTTA

AGCGCAATTAATTCTGGTAGCGTTATGCCTGTCAAG
GCGGTCCTACAACTAGGTTCGATCCTTACGACTGGA
AGATGGCTCTACACACGGACCCCCCAAACCAATTA
TAGTTACCTAGTCCTTAAAAACCATACTAGTTTGGC
TTTATTGATACTAAGACTAAGCTTACGTCCTGACTC
GCGATTAATGGACACACGTTTCCTGACAAGCTCCTC
GGGGGCCATATATATGCCTGACGCCAGAAACTGGT
CTCATTCTCGATATGAAGCGACCCAAAGCGCGGTG
TATCGTTGTCGAATCCAACTAAGATGCATCGCGCGC
GGCGGATCAATCTTACGAGACTCAGGTACTAGTGG
TATCGTGGCTGCCTTGTGACGCTTAAATCGTACTTC
GTCGCGATTGATTGTATTATAAACAATCAGCAAATT
AAATCGATGGCGGACTTTATAAAGCTAAACTACGC
CTTTAAGTTACGCGCTGTGAGCAGCTGAGGCCGGTT
CCTTAAGTTCCATACATTCTATCAATAGCGCTTCCT
GCCTAGGTATGGGCTCTAGGGCTATCTTGCTAAAGT
TGACTCAGAGAGAATTACCTCGGAATAAAACAACA
CGCGGCAGTCAGATTTTGTCACTATTTTTACGTAAC
TAGGGTGATCTCCGGAATGTCAACTCCGGGCCCCC
ACACGATGGTGGAGATCTCCTCGCCCGTGGGCTTCT
GGACTAGACGTTAGGGCATGCACATACGTTGACGA
AATTGTTACGCGGAGACGATAGAATTTATAACCTTT
CCACCATCTAGTATGAGGGATTCATACGCTGCCCTT
CTCCTAATAGGAACGTACACTAAATTAATTGCCGTG
CTACCAATGCGACTACTTTGGGATAACGGCCTGCG
GTTGTCGTCGGGTGAACTATCCTATCGTTCGACTCT
ATAGCAAGGCTTATCGTGCTAACTAATTTACATAGT
AGGACTATCGCCACACGGGATGCACATACCCGACT
ATCGGGTCCCAGAGACTACGTTGAGGAAAGCCAGG
CTTAGTTTTACACATTAACCGATGGCGTGACGGGG
ACTTTGTCGTCGGTACATAATCGTCAGGTCATCAAT
TCCTGCTGATATGGCGAAATTGCTGAGTATCTCTAT
GGACTAACAACTGCTAGGTGCTCTGGAGCCGACCG
CCGCGACATACAAGATAGACACGTCTAAACAGCTC

GTTTTCATCAACACCATCGTGCATGCCGATCGACGT
GGCACAAACAAATTGAATAGAAGGCATACTATATC
GTCTACTTGGTATGGGGCACCTTGCCGTCCAAAACC
GTTCGAAAAAAGATCTGTTTCTAATTCATCGTCAGT
CGATTTGAAATTCTCTCCCCATACGCATGGACGCAA
TAAGTATCGATTGGACACCTCCTCCCAGGTTCAATG
TGAAGTGACATCGCAACATGAACCCCGCGGGGACA
GAATGCAGTCTTCCCTGCTTAATCTCGTTGGGTACA
GCTGAAATGCAGTCAGGCGCGGATGGGGGCCCCTC
ACGGGATATGGTGATAATGTTTACTAGCTTTACACG
TTTCTAGCAGAATTGCGAAATGACGATAGCCTTCCA
CGCATATGTCCTTGCCTCTCACATCCGAATTGGCGA
TGGATGTCTCTAAATGAATTCTTATGGTCGCGACTT
TAACGCTTCCAAGATAACAACAGATGGTGCTCCTG
AATCACATCTCCTTTGATCTTGACATGGTTCCACCC
TGTTCCCCGGGCCAACCCGTTAAGCCTTACTATGTG
ATTCGACCTAATATGGATAGTCCATCCGGCCATCCG
TGTACAATAATCCACAGACTCTGTAATTTAGAATTA
CATGCACTCCTCTCATCGTATCGGCCTAATGCTAGG
ATCGGGTGCGCGATTATACGGCAACTCTGTCGATG
GCCTAGGTTGAAGGGGGATCAACACGGTGTACATA
GGCCCTACAGCTGACGTTCACGTATGATGAATGCTT
CCTCAATGTAATGCTCGAATCGAGAATTCTCAGTCT
TAAGGGCAGCCATCGGAGCACGTGGCGCGGCAATA
TTGATTATGACAGAGCTATACAGCCCACTCGGGCG
ATAGACTGCTGAGACGCAAACGTGATATTAATTAC
GATGGCTAGCATTCGACATATCATAATCAGATATTG
GGTTTAGGACCTTTATCGCAGTATTAGTACGATTTG
GTGCTGTGCGAAATCTTATGTGCGCGTGCGAAACA
ATATATTGTTCGAAGTGATATGGGATAGGTCAGTGT
CATATAATGTAAATCGGTTCGTCTGACGCGATTTAA
GGCTCACATTGTTATCGCTAATCGGGATGAACGGCT
CAAGTGCAGCATGGCACCAAGATTCCGAGGGCAAA
CGCCGCACAGTGAGGTTTGGCTCTCCCCTCTAATAT

CTTACACGTTTGTGGGATTATAGGGATCACATGGCC
ACGGCCTGTAATATTGTCATGTAGCCCGGATGATAC
CGGAATACTAAAATTGGAGGGGTTCTAGGTCATGC
TAACTGCTCGGGGCTCATGGAGTTGTAGAGTTATCA
ACAGGATCTCGGAATTCCCGTAAGCGGGATCTCCTT
GCCGATAAGTTTGTGCTGCTGCCCGTCTTCGCGCCG
GAACGCGCTTCCAAATTCTCCCTACTAACGCATGCT
GATGCACCATTGGAGCATTCTGGGATGGGCGTTTAT
CGAAACGAGTGTTTGTCTATAATGCATGACGAGGT
CTCTGCTGGGTAGAATTGGTGATTTGGAAGCGATA
CGGGTTATAGTCTCACGTACTGATGGACTAGTATGC
GTGAAGGAATCGAATACTTCGACACGATGACGTAG
GGAGCCACGCGATCAAGGACTGCCCAGTGGTCTAC
TATCTATCTTCAACAGATTGAGGGGGAGCGGTGCC
GCTGATTTAATTTTAGCATCGGTCGCTGGTTAACTT
TTAGTATCGCGCCTTTAAAGAATCTAATCTCCGTTA
GTGTCGGGTTGATTTTCTGCGAAATAGAACTAATTC
AATTGCTTATCTGCTTGATCGATTCGGAAGCCAGGG
TGGGTAGGGTAGTTACGTACGCCTGAATCTGAACC
ATCAGTCGTAATGAATTACTGAAGACGCGCGATGC
CTGGATAAAATTATCGCCTATGTCCCAACTAATGGC
ACGACAGGCTCAGAGCATGCTACTGTGTAGTGAGA
TCCGCTTATCGCCCCATTCGTGGTCGCGTTATGCCA
CTGAGTAACAAGTGATGTCCAGTGTCTAATACGAC
CGCTCGGGTCGATGGTCAAGCGGCACAGTGACATT
AACTTTTGCTTTCACATTGAACAAATTCTCCCACTT
CAGCACATGTACCCCCTGCTGCATACAGACCAGGT
CTTTTGTCCACACCTTGCACGGGTGCCTGAATGCCT
TTCCGCTGGCCTAAGCCAGTGACGTGAATGTAAAG
AGCGCTCGCACTGTAGTCATGGAGAATTATAATCG
ATAGATAAATACGTGGCGCACCACCCCAACATCCT
CGCGGGCTGTTACTAGAAATTGTGTATACCGTGGG
GGTGATTAAAAAATGGTGAGACGTGCTGTATGGTC
TTTGTGATCTCTGCTACTATTGGGTGCTGCATAAAT

CGTACCTCCAACTTGAGGCATCATAGCTACGGAAC
CCGTAAAATTGGTCATATACGCAAACACAACAGTA
AGTAGGTGGAGCCGAAGTGCTCTCGTGGCCGAAGA
CAACAACCTTTGCCCATGCCTTAAAGACTGCGTGAT
AACCGTCTTCCCATCAGGAGGTGAAGGCGATATGG
TAATCTATAGGTATTGATGGCAAGAGGTCGGAACC
CAGCTTACTCGATAGCGTTGTCGATCGCGCTTCCTG
TGCTCCTTCCTACAAAGTGGGATAGCATCATAGAC
AGGCATCCGGGTCCAATCGCCGAACGCGTCACGCA
TCGCATGATTAATTACAGTGTCGCATTACATCTAGT
ATGTATTAGGTGGGCACCGCGGTACAGCATGGACA
GGCGCTCACGGACACAAAAACGCGTCAACAAAAGT
TAGGTATGGGTGGCGCCAGGTGAAAACGCCAGCTC
TGCTATGGTCCTAAGTAATTGCAGCATGTCTTGAGA
TCTCATAGCTACCGTCTTCAGAACGATATTAGCTAA
CTTTCCCTTCCGTCTCATTACTTATGCGGGCTTCATC
GCGGTTACCGGCTGGTAAGATACGTAAGCTACACT
AGTAAGCATACTGCAGGTATGAGCCGATCCTGCAA
TTACCCATATTGGTTTTTGTATTTACACGTATGGCG
ATTACACTTCTTAAACTAGAACTCGTTTACTAATTC
TTCGTTCATACTCATGGCAATAGCATGATCTCGTAT
TACCATGTTATACGTAGTCATAGTGTGCCAACAGTA
CGTTAACCTACAATGCTCCACGCCGACCTTGTAGAA
CAGCATGATACTATATACCCGGGCATCGCGCACCG
ATAACTGCAGATCATGGAATGACCGCTCTACGTGG
ATTTAACTCGGGTGGCCCTATAGATAAATATTCTTA
CCACCGCCCTGGGATATATAGGCCGTCAGCACGTTT
ATGTCCTAGTACGCAGTACGCGCCTATTAATATAAC
AGCTGTCAGTAAGGGTCCAGAATTCTAGGGCCGAT
GAATTACAAGCAGGTGAATAGATACGATTGGGATA
TTATCACAACAACTCGCGAATGGATTATCAGTACG
AGCCACGGCCCAGCACATTATTCACCAACGGGATT
AGGTGACGCCAGTGCGTGCTGCTACTACAATGCAT
CGCGGGTGTTGACGGTTAAGGTAGCTCGGGCGCGA

TAGATGATACTGGCCCGAGACCAGTTTCTCTATATT
AACCTAGTAAGACAGGCCTGGCCCGGAAACCGTTT
CTGTACCCCGACCTAGTATAAGACTACTGGGCCGCT
AGCGGACTATTGACAAATCGCGTAGAAAATGCC
TGGGCCGTCTGCCGTCGGTTTCTTTAGCTATACCTT
GTAATTAAATACTGGACCAACCACAGTTTCTTCAGA
GTAACCTTGTACTTTAGGCCTTTACATCGTCCTCCTT
CTCCAACACGACCTTGTAGCTCACTACTGGTCCACA
GGCAGTTTCTTCAGCACCAGCTTGTATCTGATGCCT
GGTCCATTGTCCCCTTCTCCAATCGTAGCTTGTTCC
CGAATACTGGTGCTATGCCTAATTCTAGTAGATAAC
CTCGTTACCAAGCTCGTTTGCTTCAAAAGTCTCTTG
TTCCCGACGACGTAGCCAATAGCGGGCGCTCGTTC
AGTCTCTCGAGCTCTCCAGCGTTGGCCATGCCTTTC
GCTAGTCCGCCCTCTGGTCCTATACCTGGTTCCCCC
GAGCGGGGGCCAACACACACGCTGCTCTCAAAGCT
GGTTCAGGAGCGCTGGACCCTTCCAAGTCTCTAATG
CAGTCTCTAGTTGAGATTTACTGGAGCCATGCTCCC
CTCTTATGACAACTGAGGTTATGTTAGCCTGGAGCT
TAGATACCCTCTCACGCGCCCTGACGTTCTATTGTA
GTGGAACTACATTCCCGTCCCACGATAACTGACGTC
GTACTCGCGTGGAACACTAGTACCGTCCGACACCG
GCGGATGTCTTAGTTTAGTGGTACTTGTCGCCCTTC
CAACAAAAGAAGACGTCTCAATAGCGTGGTACCGT
TTTTCCGTCCTACTCTCACGGAGATCACTATGTAGT
TTCAGCGTCAGGGTGTCCTTTAAAACATAGAATCCG
TTAGGAGGTTTAGGGGCCCCCCGTCCCTCTCACGAC
GAAATAATAAATAGGGGGGAGCTCGGACCCGTCCG
TCATACCAGAGAATCTAAGGGCTGGGGGAGGATTA
GACCGTCCATCCTGTCAAAGGATGCACGTGCAGAG
GAAGAGTACACCCATCCCAGCGAAAAGTCTATCCT
CATCCTGGGGGTCCTGAAAACCATCCTCTGTCTGAG
AGTATGTTGAGGAGCGGGATGATGGCGACCCTCCC
CAACCGGGGCCCTCTGGTCCGCCTATAGTTTCAGAG

ATGAATTAGCTAAGGTTGTAGCTTATTTTCCATAGG
GTTTTGCTCCGGACCATCCGGTCGTGTAGCGCGATT
GACTTGCCGGGTTGTGTCCCCGTATCCAGGTCACGA
CCTCATGGGGAACTAGTGGCTGTCCGGCAGTATCCT
GGTACGCACCTCATGTGGTATGCGTGGCTGTTGGTC
CGTATATGGACCTATATATGGATCGAAGC

JPEG image of Indian Flag	File Size = 1981 Bytes DNA bases = 7924

In example 2, a JPEG image if Indian Flag having file size of 1981 Bytes have been encrypted in terms of DNA bases. A total of 7924 DNA bases (4-base/Byte) are required to encrypt the complete image. Since the sequence is large, fragmenting the sequence into smaller segments is required.

REFERENCES

1. Lalit M Bharadwaj*, Amol P Bhondekar, Awdhesh K. Shukla, Vijayender Bhalla and R P Bajpai. DNA-Based High-Density Memory Devices And Biomolecular Electronics At CSIO. Proc. SPIE: vol.4937, pp 319-325 (2002).

1. Clelland, C.T., Risea, V. & Bancroft, C. Hiding messages in DNA microdots. *Nature.* 399, 533-534 (1999).

2. Bancroft, et al. DNA-based steganography. U.S.Patent no. 6,312,911, November 2001.